

Fusing Meter-Resolution 4-D InSAR Point Clouds and Optical Images for Semantic Urban Infrastructure Monitoring

Yuanyuan Wang^a, Xiao Xiang Zhu^{a, b, *}, Bernhard Zeisl^c, Marc Pollefeys^c

^a Signal Processing in Earth Observation (SiPEO) Technische Universität München, Arcisstraße 21, 80333 Munich, Germany.

wang@bv.tum.de

^b Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Weßling, Germany.

xiao.zhu@dlr.de

^c Institute for Visual Computing, Department of Computer Science, ETH Zurich, CH-8092 Zurich, Switzerland.

marc.pollefeys@inf.ethz.ch

KEY WORDS: optical InSAR fusion, semantic classification, InSAR, SAR, railway monitoring, bridge monitoring

ABSTRACT:

Using synthetic aperture radar (SAR) interferometry to monitor long-term millimeter-level deformation of urban infrastructures, such as individual buildings and bridges, is an emerging and important field in remote sensing. In the state-of-the-art, deformation parameters are retrieved and monitored on a pixel-basis solely in the SAR image domain. But the inevitable side-looking imaging geometry of SAR results in undesired occlusion and layover in urban area, rendering the current method less competent for a semantic-level monitoring of different urban infrastructures.

This paper presents a framework of a semantic-level deformation monitoring by linking the precise deformation estimates of SAR interferometry and the semantic classification labels of optical images via a 3-D geometric fusion and semantic texturing. The proposed approach provides the first “SARptical” point cloud of an urban area, which is the TomoSAR point cloud textured with

* Corresponding author

attributes from optical images. This opens a new perspective of InSAR deformation monitoring. Interesting examples on bridge and railway monitoring are demonstrated.

1. INTRODUCTION

Monitoring long-term millimeter-level deformation of urban infrastructures is an important field in remote sensing. Synthetic aperture radar (SAR) interferometry is so far the only imaging-based method to assess the deformation in such accuracy. Credited to the launch of meter-resolution spaceborne SAR sensors [1], as well as the development of multipass interferometric SAR (InSAR) techniques such as persistent scatterer interferometry (PSI) and SAR tomography (TomoSAR), it is now also feasible to monitor individual buildings over a large area in high detail.

The state-of-the-art of multipass InSAR mainly leans towards the optimal parameters retrieval on a pixel-basis, i.e. the estimation of the 3-D position and the deformation parameters of scatterers. Moreover, the inevitable SAR side-looking imaging geometry results in undesired occlusion and layover especially in urban areas [2], rendering the SAR images difficult to interpret. For example, Figure 1 shows a comparison of the optical and SAR images of a same area (Potsdamer Platz, Berlin). The rich details in the optical image are buried underneath the strong reflection of the façade in the SAR image, i.e. superposition of the reflections from multiple sources in one pixel. This renders the current approach of pixel-based deformation analysis and manual identification of the region of interest less competent for understanding the semantics of the object being analyzed. Only until recently, the first investigation of semantic-level urban deformation analysis in InSAR has been presented in [3], [4].

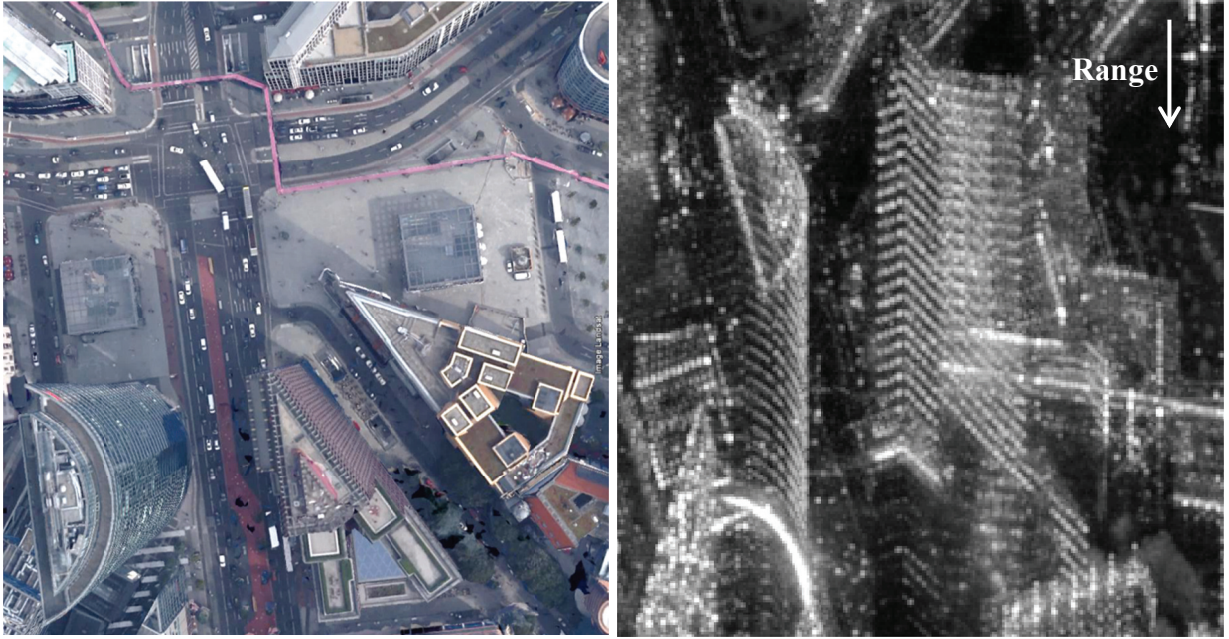


Figure 1. Comparison of optical image (left ©Google) and SAR image (right, TerraSAR-X) of the Potsdamer platz in Berlin. The rich details in the optical image are buried underneath the strong reflection of the façade in the SAR image, causing the so-called “layover”, i.e. superposition of multiple reflection sources in one pixel.

This paper proposes a framework for semantic urban infrastructure monitoring using multipass InSAR and optical images. We aim to combine InSAR and optical images by complementing the high precision deformation estimates of InSAR with the high interpretability of optical images. The focus of this paper is put on linking the optical image to the SAR image by 3-D matching and projection.

The proposed framework leads to a few contributions to the remote sensing community. We addressed the 3-D co-registration of side- and nadir-looking point clouds in urban areas, e.g. nadir-looking optical and InSAR point clouds. We also addressed the texturing of InSAR point cloud with optical semantics without explicit reconstruction of 3-D model. Last but not least, the joint deformation analysis using optical and SAR images leads to new perspective in multipass InSAR.

2. METHDOLOGY OVERVIEW

The general framework of the proposed approach is shown in Figure 2. The framework applies to a stack of SAR images and a pair of (or more) optical images. The focus of this paper is put on linking the attributes from optical image to the SAR image by 3-D matching and projection. The basic idea is to match the 3-D models derived from SAR and optical images respectively. As a result, the 2-D SAR and optical images will also be matched. Only then, the following semantic label texturing and joint deformation analysis can be conducted. The detailed procedures of the proposed framework are as follows.

1. 3-D reconstructions

- a. Retrieve the 3-D positions and deformation parameters of the scatterers from the SAR image stacks. Since urban area is of our main interest, tomographic SAR inversion (TomoSAR), including SAR tomography and differential SAR tomography, is employed in order to resolve a substantial amount of layovered scatterers.

TomoSAR is the most computationally expensive step in the framework. In addition, TomoSAR and other multipass InSAR algorithms typically requires a fairly large SAR image stack (>20 images). The computational and image resource are the main limitation for this step.

- b. Retrieve the 3-D positions of points from the optical images using stereo matching with structure from motion (SfM) if necessary. For covering large urban area, aerial or spaceborne images are preferred. This step also calibrates the camera parameters.

Stereo matching and SfM are well studied topics. Many matured algorithms and software are readily available.

2. 3-D matching: Co-register the TomoSAR point cloud and the optical point cloud.

The main challenges present in this step are the different modalities of optical and TomoSAR

point clouds, i.e. nadir-looking and side-looking, as well as the relatively large anisotropic noise in the TomoSAR point cloud. However, considering the large amount of points compared to the few co-registration parameters to be estimated, the co-registration accuracy is expected to be high enough for the following steps.

If only a single optical image is available, which means no optical point cloud can be reconstructed, one can try to match higher level features of the 2-D optical image with the 3-D InSAR point cloud, e.g. matching 2-D and 3-D lines [5]. However, this has not been implemented in the proposed framework.

3. **Optical image classification:** applying semantic classification to the optical images.

This part is not the focus of this paper. Depending on the application, different classification algorithms can be applied.

4. **Semantic texturing:** Texture the InSAR point cloud with the attributes derived from optical images, e.g. RGB color, semantic classification label, etc.

The main challenge of this step is to project the optical image to TomoSAR point cloud without explicit 3-D surface reconstruction in the TomoSAR point cloud. Therefore, we choose point-based rendering technique.

The main limitation of this step is the relatively poor positioning accuracy (1 to 10m) of spaceborne TomoSAR point cloud. This error will directly translate to the projection accuracy of the TomoSAR points in optical image.

5. **Deformation analysis:** Perform further analysis on semantic-level in the InSAR point cloud based on their semantic class.

The main bottleneck of this step remains at the positioning accuracy of TomoSAR points, which limits the joint deformation analysis to be at an object level instead of at a point level.

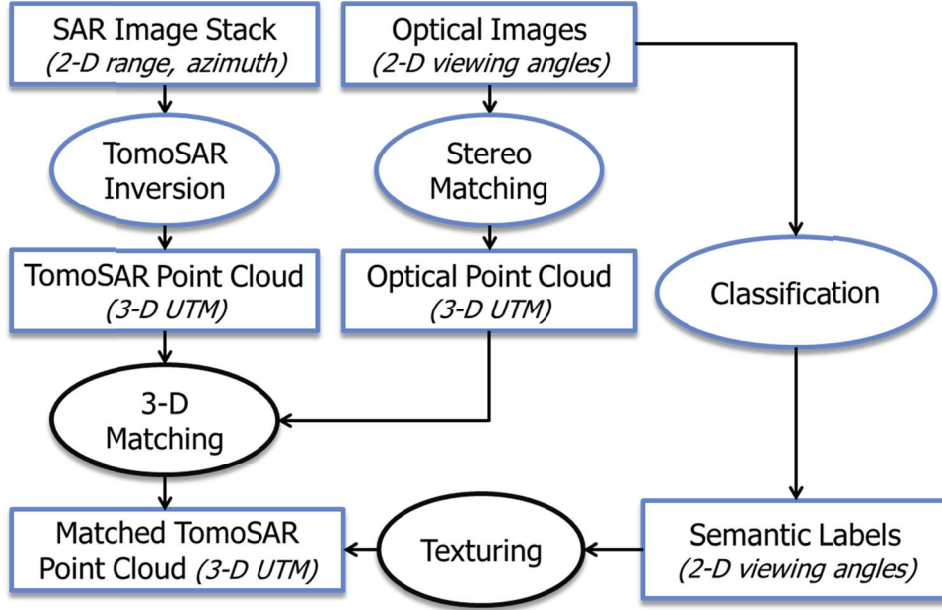


Figure 2. Flowchart of the proposed framework. It matches the 3-D TomoSAR and optical point clouds, and then transfers the semantic classification label from optical images to the TomoSAR point cloud via a point-based texturing. This paper focuses on the 3-D matching and texturing which are shown in black ellipses. The coordinate system of each dataset in the flowchart is indicated by the italic text in the bracket.

3. RELATED WORK

3.1 Differential SAR Tomography

Regarding urban area monitoring using multipass InSAR, TomoSAR in its differential formulation [6]–[9] is the most competent method because of its capability of separating layovered scatterers in one pixel and estimating their motion parameters. Figure 3 shows the SAR imaging geometry a range-azimuth pixel, where the reflections of multiple sources are layovered in one range-azimuth-elevation resolution cell.

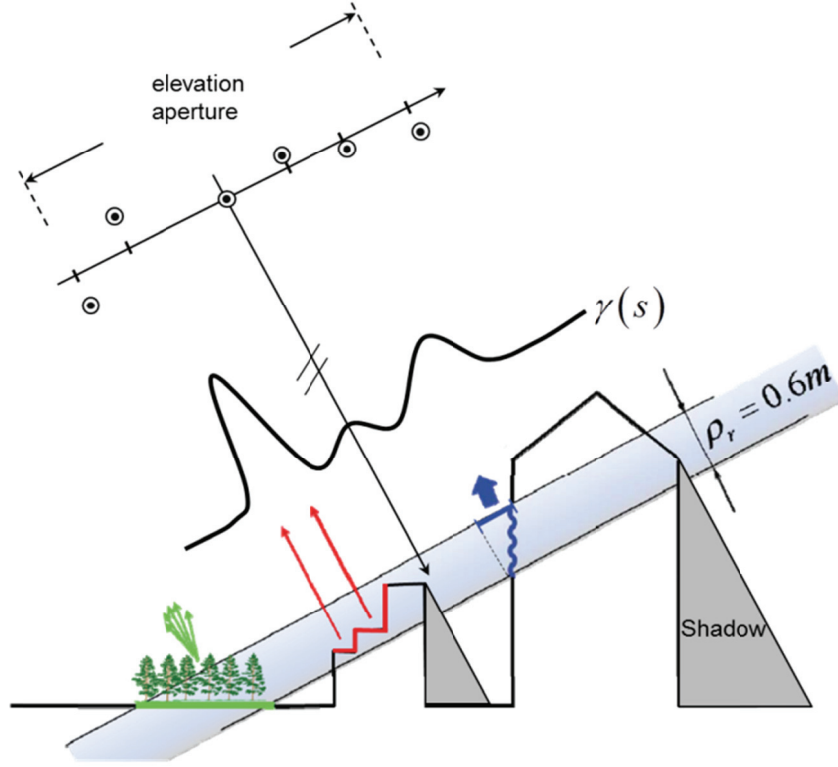


Figure 3. SAR imaging geometry (modified after [8]). Multiple scatterers are layovered in one resolution cell, which became one range-azimuth pixel. TomoSAR retrieves the whole reflectivity profile $\gamma(s)$ using multiple images acquired at slightly different antenna positions.

The reflectivity profile $\gamma(s)$ along the third dimension *elevation* s is integrated into one range-azimuth pixel, as shown by the following equation [9]–[11]:

$$g_n = \int_{\Delta s} \gamma(s) \exp(-j2\pi\xi_n s) ds \quad (1)$$

where g_n is the complex SAR pixel value in the n th image, Δs is the extent of the elevation direction, and $\xi_n = 2b_n/(\lambda R)$ is the spatial frequency which is related to the baseline b_n of the n th image, the radar carrier wavelength λ , and the radar range R .

TomoSAR retrieves the full reflectivity profile $\gamma(s)$ from multiple observations. Equation (1) can be discretized as

$$\mathbf{g} = \mathbf{R}\mathbf{y} + \mathbf{\varepsilon} \quad (2)$$

where $\mathbf{g} \in \mathbb{C}^N$ is the observation vector, $\mathbf{R} \in \mathbb{C}^{N \times L}$ is the so-called steering matrix which is a discrete Fourier transform according to ξ_n and the discretization level L in the elevation direction, and $\mathbf{\varepsilon} \in \mathbb{C}^N$ is the noise assumed to be complex circular Gaussian distributed. The discretization level L is usually much larger than the number of images N . Therefore, equation (2) is often underdetermined, and can be even ill-posed due to the irregularity and small span of the baseline. Hence, regularization is required. The Tikhonov regularization of the estimator is equivalent to the maximum a posteriori (MAP) estimator with white noise and white prior, which is usually realized using singular value decomposition-based methods [8], [10].

From the reconstructed reflectivity profile, multiple scatterers can be separated, and hence the full 3-D (range, azimuth, and elevation) reflectivity distribution is obtained. The motion of scatterers can also be modelled and estimated by introducing new dimensions into the TomoSAR system model attributing to the motion of the scatterers [6]. For a motion model with M parameters, the extended TomoSAR system model is a $M+1$ dimensional discrete Fourier transform [12]:

$$\begin{aligned} g_n = \int_{\Delta p_M} \cdots \int_{\Delta p_1} \int_{\Delta s} \gamma(s) \delta(p_1 - P_1(s), \dots, p_M - P_M(s)) \\ \exp\left(-j2\pi\left(\xi_n s + \eta_{1,n} p_1 + \cdots + \eta_{M,n} p_M\right)\right) ds dp_1 \cdots dp_M \end{aligned} \quad (3)$$

where p_1, \dots, p_M are the M motion dimensions, and $P_1(s), \dots, P_M(s)$ are the motion dimensions as functions of the elevation s . $\eta_{1,n}, \dots, \eta_{M,n}$ are the M motion models, e.g. $\eta_n = 2t_n/\lambda$ for linear motion model and $\eta_n = 2\sin(2\pi(t_n - t_0))/\lambda$ for seasonal motion model, with t_n being the acquisition time in unit of year of the n th image w.r.t. the master image, and t_0 the initial offset also in unit of year.

3.2 Multi-view 3-D Reconstruction from Optical Images

Dense 3-D reconstruction from optical images is a classical topic in photogrammetry that is usually carried out by stereo matching. On the other hand, many algorithms were developed over the last decade to exploit the available unstructured image data. At the core of all algorithms are local image features which are matched to each other among the set of captured images. These relations allow recovering both, the camera motion (camera parameters) and the scene structure (3-D geometry), which is why these algorithms are referred to as *Structure-from-Motion* (SfM) [13]–[17]. Tremendous progress has been made and led to impressive results, even modelling whole cities from optical images only [13]–[15]. In the state-of-the-art, SfM is typically followed by a dense reconstructions step, e.g. semi-global matching [18], for 3-D reconstruction of urban area.

The principle of SfM is shown in Figure 4. If all the red dots in the image are known to correspond to the same 3-D point, given its 3-D position and the cameras parameters, one can predict its positions in the images. Consequently, one can estimate its 3-D position and the cameras parameters by adjusting their values so that the difference between the predicted positions of the 3-D points in the images w.r.t. the observed ones is minimized. For example, Camera C has a large reprojection error; hence, the 3-D point estimates and the camera parameters should be adjusted accordingly.

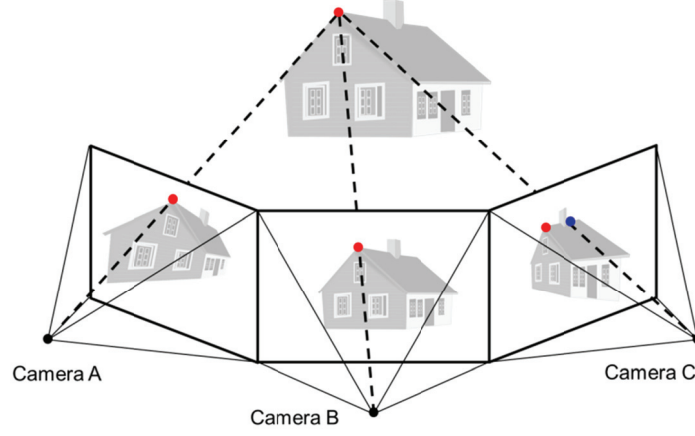


Figure 4. 2-D projections of a 3-D house in three different camera positions and orientations. The red dots are a known correspondence point between the three images. SfM estimates the unknown 3-D position of the point and the camera parameters which minimize the reprojection error (the predicted point location in image w.r.t the observed one, e.g. the blue point in Image C is the predicted location. Camera C has a large reprojection error.)

Therefore, the SfM system model can be formulated as follows. Given a set of corresponding image points, where we use $\tilde{\mathbf{x}}_{ij}$ to denote the i th points in image j , the projection from their real 3-D coordinates \mathbf{x}_i is [13]

$$\tilde{\mathbf{x}}_{ij} = f_j \Pi(\mathbf{R}_j(\mathbf{x}_i - \mathbf{t}_j)) \quad (4)$$

where f is the scalar camera focal length in number of pixels, \mathbf{R} , \mathbf{t} are the camera orientation matrix and translation vector respectively, and Π is the projection function: $\Pi(x, y, z) = (x/z, y/z)$

. SfM is to estimate the unknown 3-D position and the camera parameters from the observations $\tilde{\mathbf{x}}_{ij}$.

Assuming a Gaussian distributed reprojection noise, the maximum likelihood estimator of the parameters is [13]

$$\{\hat{\mathbf{x}}_i, \hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j, \hat{f}_j\} = \arg \min_{\mathbf{x}_i, \mathbf{R}_j, \mathbf{t}_j, f_j} \sum_{i,j} \left\| \tilde{\mathbf{x}}_{ij} - f_j \Pi(\mathbf{R}_j(\mathbf{x}_i - \mathbf{t}_j)) \right\|_2^2 \quad (5)$$

Sometimes, the camera radial and tangential distortions should also be taken into account, and should be estimated [16], [19].

3.3 Optical and SAR Images Co-registration

Geometric co-registration is the basis of many higher level fusion algorithms. The current techniques are mostly based on matching of 2-D spatial features, e.g. contours, edges, templates, etc. [20]–[25]. They are good for large scale analysis, but not applicable for the analysis of individual building, especially for high resolution data.

Therefore, for the co-registration of optical and SAR images of urban area including many studies of their joint analysis, a precise 3-D model is crucial. For example, [26] used 3-D model reconstructed from photogrammetric measurements; [27] used precise airborne light detection and ranging (LiDAR) point cloud; and [28] used PS point cloud retrieved from tens of high resolution TerraSAR-X spotlight images. Without a precise 3-D model, one needs to assume a strict prior [29], such as rectangle footprint etc.

Considering the complex 3-D topography of urban area, there has not been an effective algorithm of matching SAR and optical image yet. The method presented in this paper transform the 2-D images matching to 3-D point clouds co-registration. Handful of papers have explained the co-registration of very high resolution InSAR point clouds [30], [31]. In a broader context of 3-D point clouds co-registration, one of the most successfully methods is the iterative closest point (ICP) [32]. As its name implies, ICP solves the following equation at each iteration:

$$\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\} = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_i \|\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}\|_2^2 \quad (6)$$

where \mathbf{R} and \mathbf{t} are the rotation matrix and the translation vector, and \mathbf{x}_i and \mathbf{p}_i are the respective corresponding points in the reference (optical) and target point clouds (TomoSAR). Considering the

anisotropic noises in the point clouds, [33], [34] take into account the covariance matrix $\mathbf{C}_{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i}$ of the residual $\boldsymbol{\varepsilon}_i = \mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}$:

$$\{\hat{\mathbf{R}}, \hat{\mathbf{t}}\} = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_i (\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t})^T \mathbf{C}_{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i}^{-1} (\mathbf{x}_i - \mathbf{R}\mathbf{p}_i - \mathbf{t}) \quad (7)$$

4. CO-REGISTRATION OF TOMOSAR AND OPTICAL POINT CLOUDS

The 2-D SAR and optical image matching is transferred to 3-D point clouds matching. Any InSAR methods, including TomoSAR, deliver relative estimates with respect to a reference point whose height is usually unknown. Such differential measurement is always performed in multipass InSAR processing in order to mitigate some common error in each image, e.g. atmospheric effect or orbit errors. As a consequence, while being geo-coded into Universal Transverse Mercator (UTM) coordinate, the TomoSAR point cloud is shifted from its true position by an unknown amount due to the unknown height of the reference point. The co-registration is then the estimation of the translation between two rigid point clouds, subject to a certain tolerance on rotation and scaling. A good matching is the key to find the camera position w.r.t. the SAR image.

Considering the different image characteristics of side-looking SAR and nadir-looking aerial images used in our experiment, such co-registration problem can be generalized to the matching of side-looking and nadir-looking point clouds. The following issues must be taken into account:

- Due to the nadir-looking geometry of the optical images, façade points barely appear in the optical point cloud while they are prominent in the TomoSAR point cloud. This difference is exemplified in Figure 5, where the left and the right subfigures correspond to the TomoSAR and the optical point clouds of a same area in Berlin, respectively.
- The noise of the TomoSAR point cloud is extremely anisotropic. The Cramér–Rao lower bound (CRLB) of the elevation estimate for pixels of high signal-to-noise-ratio is [1], [8]

$$\sigma_{\hat{s}} = \frac{\lambda R}{4\pi\sigma_b\sqrt{N}\sqrt{2SNR}} \quad (8)$$

where σ_b is the standard deviation of the baseline, N is the number of images, and SNR is the signal-to-noise-ratio in linear scale. For a typical stack of 50 TerraSAR-X images, the CRLB of a 10dB pixel is about 1m which represents almost the best situation. In the contrary, the range and azimuth positions can be as accurate as 1 cm [35]–[37]. Considering the high dynamic range of the SNR in the scene, the elevation estimates are at least *one to two orders of magnitude* worse than those of range and azimuth. In UTM coordinate, the error in elevation is mostly transferred to the east and up directions for TerraSAR-X data, since the satellite is roughly on a polar orbit.

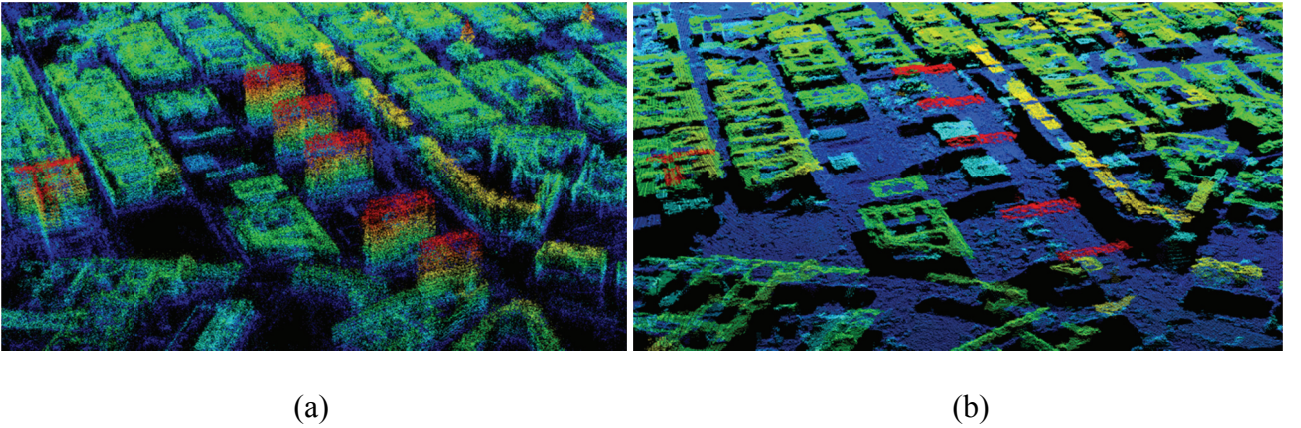


Figure 5. (a) TomoSAR point cloud of high-rise buildings, and (b) the optical point cloud of the same area. Points are color-coded by height. Building façades are almost invisible in the optical point cloud, while they are prominent in the TomoSAR point cloud. The coordinate is UTM.

4.1 Co-registration algorithm

Typical point cloud co-registration problem employs ICP algorithm. However, ICP requires a good initial alignment. The unique modalities of optical and TomoSAR point clouds have driven our algorithm to be developed in the following way:

1 Edge extraction

- a. The optical point cloud is rasterized into a 2-D height image.
- b. The point density of TomoSAR point cloud is estimated on the rasterized 2-D grid.
- c. The edges in the optical height image and in the TomoSAR point density image are detected. These edges both correspond to the façade locations in the point clouds.

2 Initial alignment

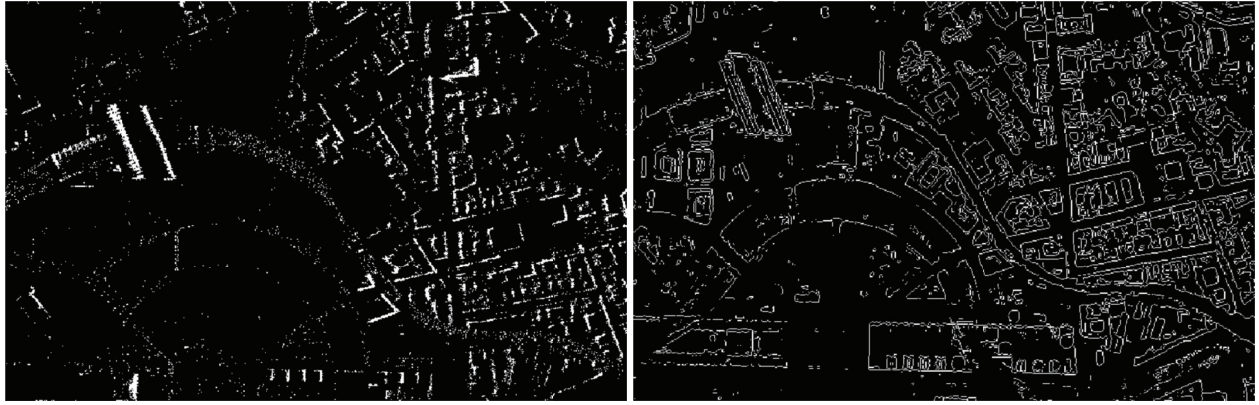
- a. Horizontally by cross-correlating the two edge images.
- b. Vertically by cross-correlating the height histogram of the two point clouds.

3 Refined solution

- a. The façade points in the TomoSAR point clouds are removed, because the optical point cloud contains nearly no façade point.
- b. To refine the co-registration of the two point clouds, an *anisotropic* ICP (AICP) with robustly estimated covariance matrices was applied. The covariance is estimated using a robust M-estimator.

2-D Edge Extraction

The rasterized optical height image and TomoSAR point density image are produced by computing the mean height and counting the number of points, inside each grid cell, respectively. The edges can be extracted from these two images using an edge detector, such as Sobel filter [38]. The thresholds in the edge detector are decided adaptively, so that the numbers of edge pixels in the two edge images are on the same order. Figure 6 is a comparison of the binary edge images of TomoSAR and optical point clouds. These edges correspond to the position of façades.



(a)

(b)

Figure 6. (a) An example of the binary edge image of the TomoSAR point cloud in downtown Berlin, and (b) the edge image of the optical point cloud at the same area.

Coarse Alignment

The coarse alignment provides an initial solution to the ICP algorithm which is known to suffer from finding a local minimum. The coarse alignment in the horizontal and the vertical directions are carried out independently. The horizontal shift is found by cross-correlating the binary edge images of the two point clouds. The larger the area, the more prominent and unambiguous the correlation peak will be. Figure 7 shows the 2-D correlation of two edge images, where a single prominent peak is found. The vertical shift is found by cross-correlating the height histograms of the two point clouds, which is shown in Figure 8. The accuracy of the coarse shift is of course limited by the discretization level in the edge image and the height histogram. Nevertheless, this is sufficient for our application.

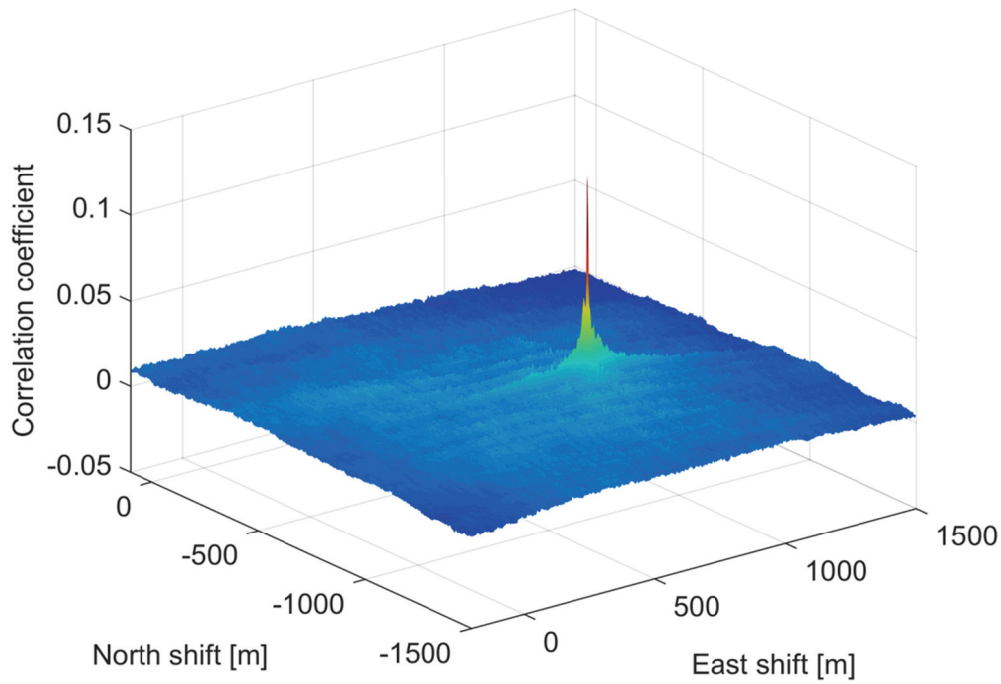


Figure 7. 2D cross-correlation of the edge images of TomoSAR and optical point clouds. The larger the area of the edge images, the more prominent and unambiguous the correlation peak will be. For the size of an entire city, a single prominent peak can usually be found.

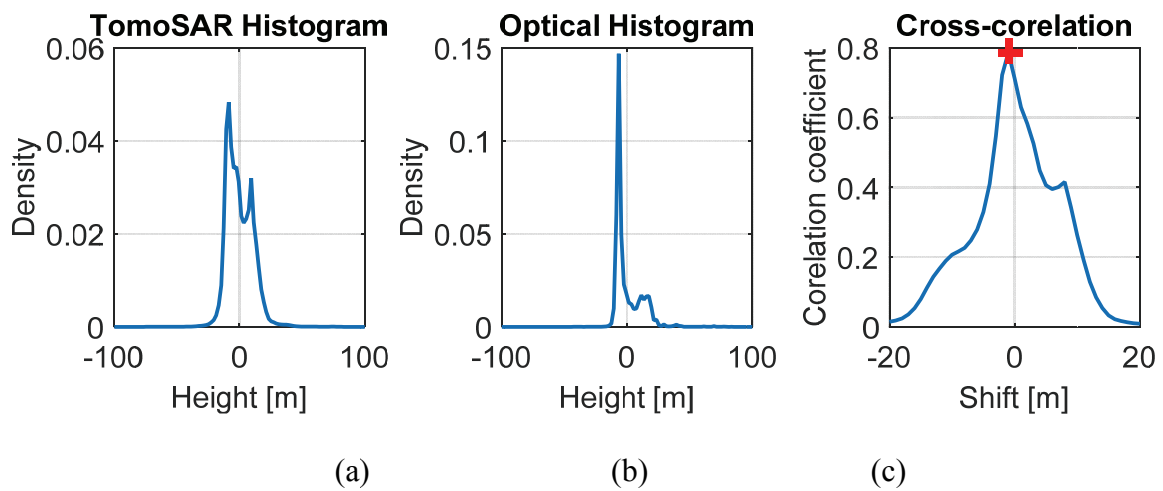


Figure 8. (a) The height histogram of the TomoSAR point cloud, (b) the height histogram of the optical point cloud, and (c) the correlation of (a) and (b), where the red cross marks the peak position indicating the vertical shift.

Robust Anisotropic ICP

The final solution is obtained using an anisotropic ICP algorithm based on the coarse alignment. The key of the anisotropic ICP lies on the covariance matrix $\mathbf{C}_{\mathbf{e}_i \mathbf{e}_i}$ of each point pair \mathbf{x}_i (optical) and \mathbf{p}_i (TomoSAR). Assuming the noise of optical point cloud is not correlated with that of the TomoSAR point cloud, $\mathbf{C}_{\mathbf{e}_i \mathbf{e}_i}$ can be computed by [33]

$$\mathbf{C}_{\mathbf{e}_i \mathbf{e}_i} = \mathbf{R} \mathbf{C}_{\mathbf{p}_i \mathbf{p}_i} \mathbf{R}^T + \mathbf{C}_{\mathbf{x}_i \mathbf{x}_i} \quad (9)$$

where $\mathbf{C}_{\mathbf{p}_i \mathbf{p}_i}$ and $\mathbf{C}_{\mathbf{x}_i \mathbf{x}_i}$ are the covariance matrix of the TomoSAR and the optical points, respectively.

As mentioned before, the dynamic range of the SNR of TomoSAR points is very high. In particular, closely space pixels may have dramatic SNR difference. This renders their positioning error distribution non-ergodic when selected for covariance estimation. To this end, a robust estimation of the covariance matrix — a topic has not been addressed in any previous ICP algorithms — is necessary.

To this end, we introduce an M-estimator [39] to robustly estimate the covariance matrix. Given a set of M samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots] \in \mathbb{R}$, M-estimator is an iteratively re-weighted sample covariance matrix [40]–[42],

$$\hat{\mathbf{C}}_k = \frac{1}{M} \sum_{i=1}^M w(\mathbf{e}_i^T \hat{\mathbf{C}}_{k-1}^{-1} \mathbf{e}_i) \mathbf{e}_i \mathbf{e}_i^T \quad (10)$$

where k is the iteration index, $\mathbf{e}_i = \mathbf{x}_i - \hat{\mathbf{x}}$ with $\hat{\mathbf{x}}$ being the estimated sample mean, and $w(x)$ is a robust weighting function. The weighting function downweights heavily deviated samples whose whitened residual $\mathbf{e}_i^T \hat{\mathbf{C}}_k^{-1} \mathbf{e}_i$ is large. We choose the t-distribution weighting function which can handle heavily tailed distributions [41]:

$$w(x) = \frac{3 + \nu}{\nu + x} \quad (11)$$

where $\nu \in \mathbb{R}^+$ is the degree of freedom (DoF) of the t-distribution. The lower the ν , the more robust but less efficient (under Gaussian distributed noise) of the estimator, and *vice versa*. ν is typically set to 2~5.

5. TOMOSAR POINT CLOUD TEXTURING

Texturing the TomoSAR point cloud is the step to link the semantics from the optical images to the deformation parameters of the TomoSAR point cloud, which can only be done upon obtaining the position and the orientation of the cameras w.r.t. the TomoSAR point cloud. This is equivalent to solving the visibility/occlusion problem given a camera viewpoint. Figure 9 illustrates the occlusion of TomoSAR points where certain points are only visible to certain images.

For solving the occlusion in a spaceborne TomoSAR point cloud, the following two issues should be taken into consideration.

- Typically, solving the occlusion of a point cloud requires reconstructing the 3-D surface and testing the depth of each point. This is known as the z-buffer algorithm. But reconstructing a reliable 3-D surface from TomoSAR point cloud still remains a challenging task, although there are some pioneer works on reconstructing façades and roofs [43]–[45]. Therefore, the employed technique should avoid surface reconstruction.
- Secondly, one point can be visible from multiple images. Strategy should be developed to choose the appropriate textures from multiple optical images. The textured point cloud should be visually consistent.

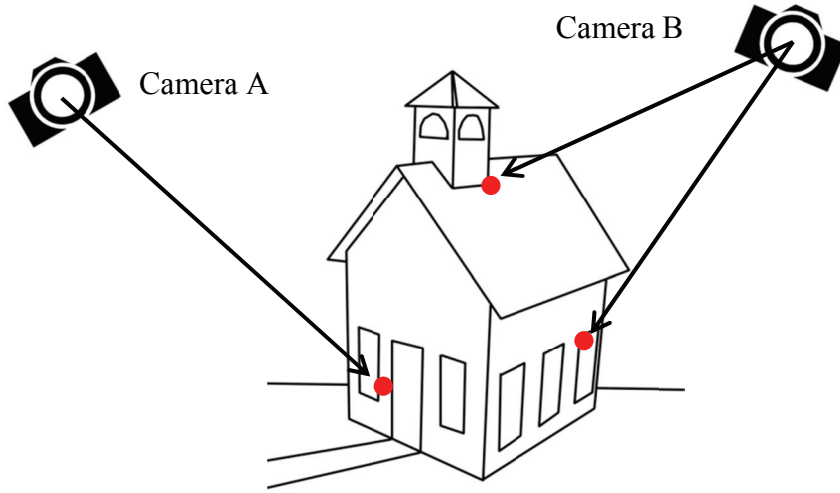


Figure 9. Illustration of TomoSAR point cloud texturing. The red dots represent TomoSAR points. The leftmost point is only visible to Camera A, whereas the right ones are only visible to Camera B.

5.1 Point-based Rendering

Hidden point removal (HPR) operator was introduced in [46], which is an elegant solution to avoid reconstructing surface or even surface normal in solving the occlusion problem. HPR inverts the point cloud using certain functions, so that all the visible points will lie on the convex hull of the inverted point cloud. And hence, finding the visible points becomes equivalent to compute the convex hull of the inverted point cloud.

The most common inverting function is the *spherical flipping*. Given a point \mathbf{x} , the spherical flipping function is defined as follows [46]:

$$\hat{\mathbf{x}} = f(\mathbf{x}) = \mathbf{x} + 2(r - \|\mathbf{x}\|) \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (12)$$

where r is radius of a multi-dimensional sphere. Figure 10 illustrates the transformation function and HPR in 2-D. The blue dots are the original noisy point cloud. And the red dots are the inverted

one given the camera viewpoint O . The bottom of the blue point cloud is visible to the camera. After inversion, these points lie closely to the convex hull of the inverted point cloud.

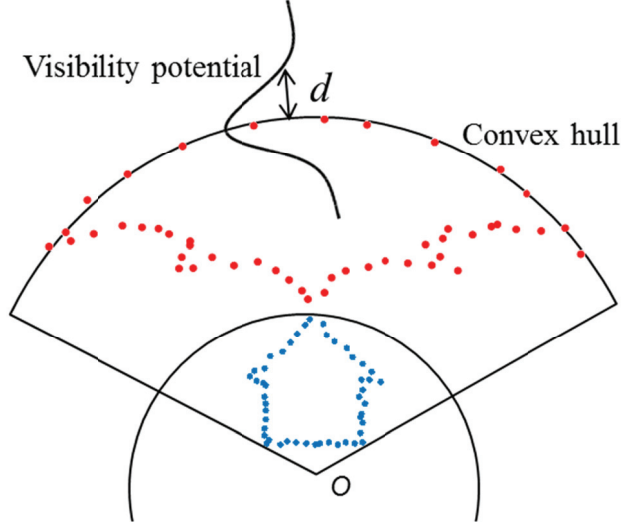


Figure 10. The improved hidden point removal operator. The blue dots are the original point cloud. Some noise was added to indicate the noise of the input TomoSAR point cloud. And the red dots are the inverted point cloud given the camera viewpoint O . The visible points will lie closely to the convex hull. Considering the noise of the point cloud, the visibility potential gives a soft threshold to the point visibility. d is the point to convex hull distance.

However, as mentioned in [47], the original HPR performs poorly in presence of noise. Considering the much larger noise in the spaceborne TomoSAR point cloud than in a typical computer vision problem, the HPR is not directly applicable.

[47] suggested to include a buffer zone around the convex hull, so that all the point in the buffer zone can be marked as visible. Inspired by this, we introduced a *visibility potential* which is a Gaussian function of the point-to-convex hull distance d :

$$p(\hat{\mathbf{x}}) = \exp\left(-\frac{d^2}{2\sigma_{\hat{\mathbf{x}}}^2}\right) \quad (13)$$

where $\sigma_{\hat{\mathbf{x}}}$ is the norm of the total noise standard deviations (all three directions) in the inverted point cloud. $\sigma_{\hat{\mathbf{x}}}$ is related to original point cloud. Consider the very anisotropic noise of TomoSAR point cloud, we define the standard deviation in the native SAR coordinate range, azimuth, and elevation being $[\sigma \quad \sigma \quad k\sigma]$, where $k > 1$. As mentioned in Chapter 4, this factor is about 10 to 100. Hence, the noise in range and azimuth can be neglected. Transforming to UTM coordinate, the standard deviation for east (x), north (y), and up (z) is approximately $[k\sigma/\sqrt{2} \quad \sigma \quad k\sigma/\sqrt{2}]$ (assuming a near polar orbit, and 45° of incidence angle). The maximum $\sigma_{\hat{\mathbf{x}}}$ is derived to be:

$$\sigma_{\hat{\mathbf{x}}} = \frac{4rk\sigma}{\|\mathbf{x}\|_{\min} - k\sigma} - k\sigma \quad (14)$$

where $\|\mathbf{x}\|_{\min}$ is the minimum distance of the original point cloud to the viewpoint.

5.2 Consistency Check

Many points are visible to multiple images. Selecting the most appropriate image for texturing should consider the visual consistency of the textured point cloud, as well as the optical quality of the texture, since pixels located far away from the center of an image often have oblique angle with poor resolution.

We use the information of the neighbouring points to check the consistency of the texturing, so that closely spaced points will select texture from a same optical image. Firstly, for each point, a score similar to [5] is obtained w.r.t. each image as follows:

$$c_j = p_j d_j \quad (15)$$

where j is the image index, p_j is the visibility potential (equation (13)), and d_j is the inverse of the distance to the viewpoint of camera j . d_j can be regarded as the resolution of the image at the point.

Equation (15) basically favors image with greater visibility and shorter distance. The image with the largest score is selected as the initial texture source image.

Secondly, to decide the final texture source image for each point, we search the k nearest neighbors of each point, and perform a weighted multi-class majority voting. The weight is a Gaussian kernel as a function of the distance of the neighbor points to the center point. Largest weight is assigned to the point itself, and lower weights are assigned to the points further away. The image with the highest votes is selected as the final texture source.

6. PRACTICAL DEMONSTRATION

This section demonstrates the proposed framework by examples of railway and bridge monitoring, which are important ground infrastructures that require systematic monitoring. Monitoring large façades is also a potential application, given oblique aerial images which are unfortunately not available in this study.

6.1 Datasets

SAR

The SAR datasets consist of two stacks of TerraSAR-X high resolution spotlight images of Berlin with spatial resolution of about 1m. The two image stacks were acquired from Feb. 2008 to Mar. 2013 with about 100 images each. They were acquired at ascending and descending orbits respectively, which provide a full coverage of the whole city. More information of the two SAR image stacks can be found in Table 1.

Table 1. Information of the SAR image stacks

Beam	Orbit	Incidence angle	No. of images	Temporal span	Baseline span
42	Descending	36°	109	5.3 year	363m
57	Ascending	42°	102	5.3 year	840m

Optical

The optical images are 9 UltraCam aerial images of Berlin acquired in March 2014 at an altitude of about 4000 m. The ground spacing is approximately 20cm per pixel. The camera positions and orientations are measured by onboard GPS and inertial measurement unit. After SfM adjustment, their standard deviations are about 5cm and 5×10^{-4} degree, respectively.

6.2 3-D Reconstructions

TomoSAR

The InSAR stacking was performed by the German Aerospace Center (DLR)'s *IWAP* processor—the integrated wide area processor [48] based on its predecessor *PSI-GENESIS* [49]. The atmospheric phase correction and the following D-TomoSAR processing was performed by *Tomo-GENESIS* [50], [51] – a TomoSAR software of the DLR for large urban areas monitoring. It is developed based on the work of [8], [9], [12], [51]–[53]. It delivers a 5-D point cloud – 3-D position plus linear deformation rate and amplitude of seasonal motion - from a stack of SAR images of urban area.

Figure 11 is the TomoSAR point clouds from the two image stacks reconstructed by the SVD-Wiener [8] (Fourier-based) algorithm in the Tomo-GENESIS. The two point clouds are co-registered using a feature-based matching algorithm which estimates and matches common building edges in the two point clouds [31]. The combined TomoSAR point cloud has about 40 million points. The point density is about $10^6/\text{km}^2$ which is even comparable to some LiDAR product. It enables the retrieval of the features of individual building. Behind each of these points, the linear deformation rate and the amplitude of seasonal deformation were also estimated.

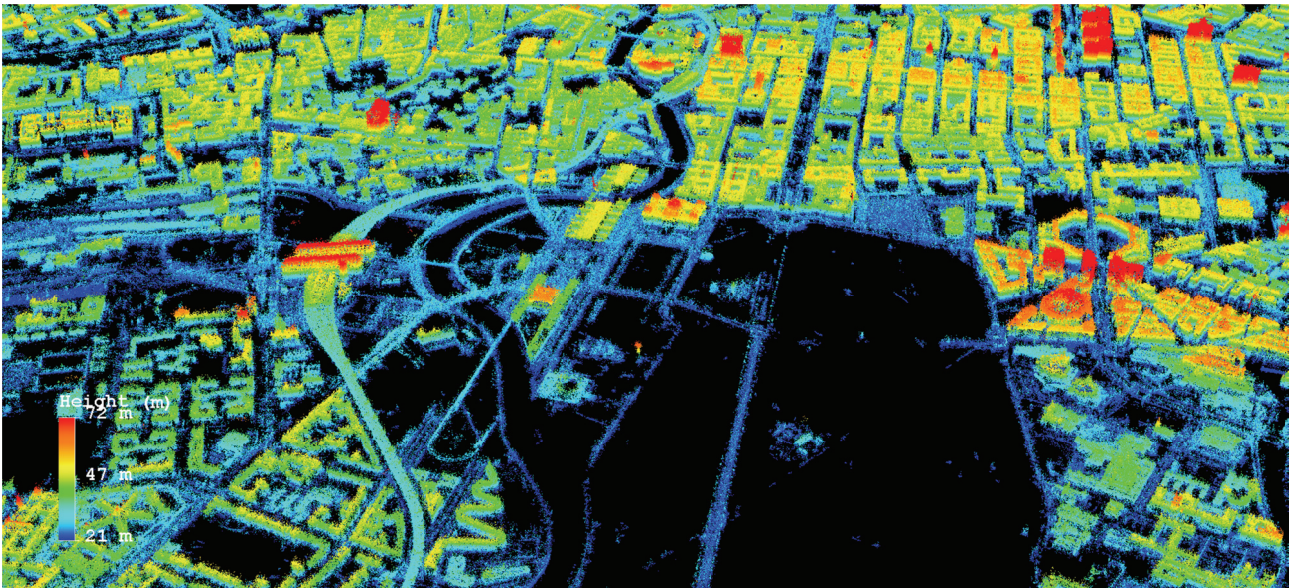


Figure 11. The fused TomoSAR point cloud of Berlin, which combines the result from an ascending stack and a descending stack. The height is color-coded. The coordinate is UTM.

Optical

The 3-D reconstruction in optical images was carried out by the commercial software Pix4D [54]. It calibrates the camera model, as well as the positions and orientations of the cameras. Figure 12 is the reconstructed 3-D optical point cloud showing the identical area as that in Figure 11. The point cloud contains 35 million points. However, basically no façade point exist which contradicts the rich façade points in the TomoSAR point cloud.

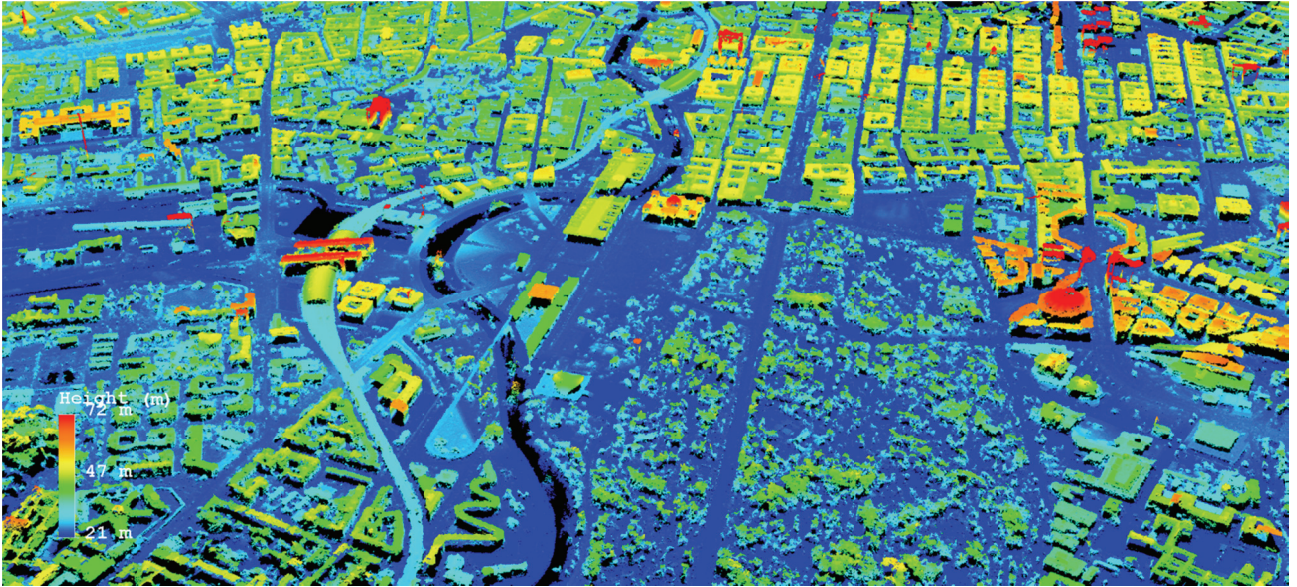


Figure 12. Optical point cloud of Berlin reconstructed using structure from motion and dense matching. The height is color-coded. The colorbar and area are identical to Figure 11. The coordinate is UTM.

6.3 Point Clouds Accuracies

TomoSAR point cloud

InSAR methods, like TomoSAR, always provide *relative estimates*, i.e. positioning and deformation estimates are referred to a selected reference point. The relative positioning accuracy, i.e., relative to the reference point, depends on the SNR and the number of images (equation (8)). For the TomoSAR point clouds in this paper, the accuracy ranges from 4m to 4cm, which correspond to SNR of -10dB to 10dB.

Absolute TomoSAR/InSAR, also referred to as geodetic TomoSAR/InSAR, is still a very new and open research topic [55]. Preliminary study on several manually selected points with an extremely high SNR suggests a positioning accuracy of up to 20cm [55], [56].

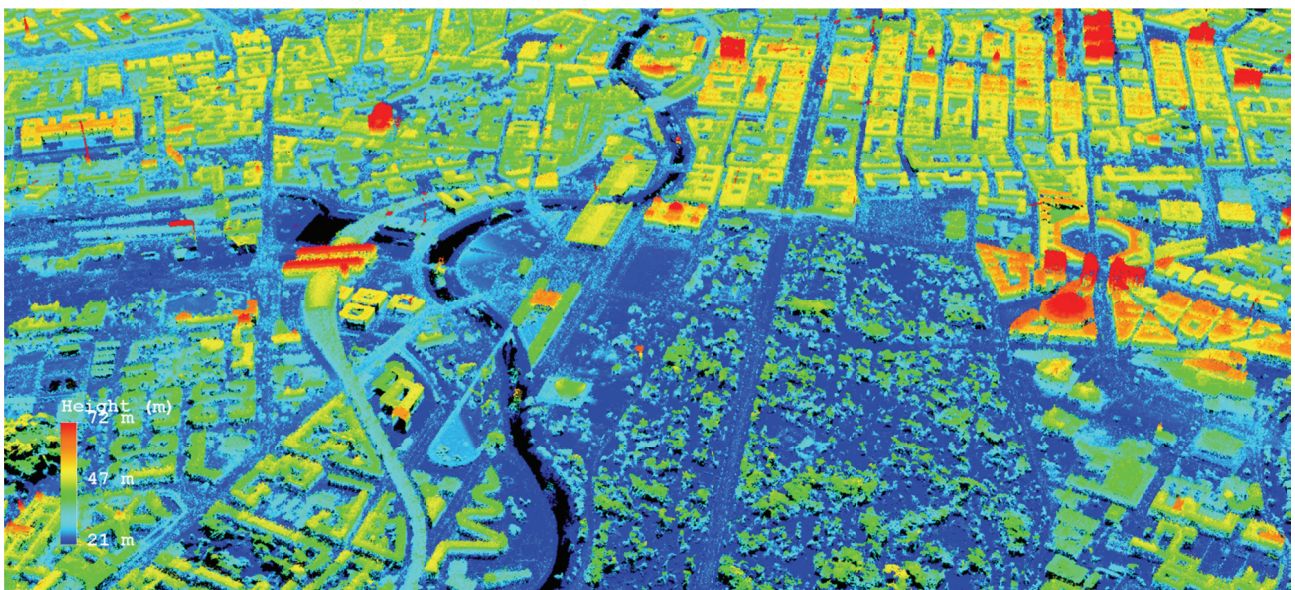
Optical point cloud

The global localization accuracy of the optical point cloud is given by the accuracy of the exterior parameters of the camera, which are about 5cm in position and 5×10^{-4} degree in orientation after adjustment. This translates to about 10cm at 4000m altitude where the images were taken.

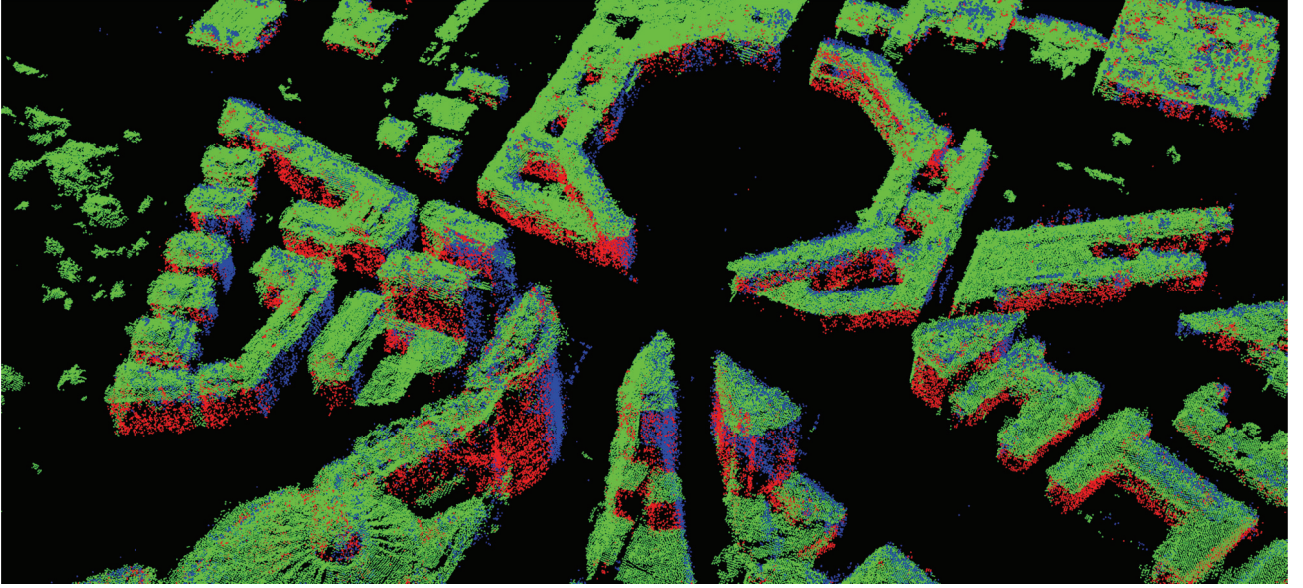
The relative accuracy of the optical point cloud depends on the stereo matching algorithm, e.g. semi-global matching. Since we used a commercial software, the algorithm behind it is unknown. But at the best, the standard deviation of the 84 ground control points is 3cm. Also by checking some flat surfaces in the optical point cloud, the accuracy in height is about 20cm.

6.4 Point Clouds Co-registration

Since the optical point cloud is quite accurate, and ground control points were used, the scaling and rotation are therefore not considered in the co-registration. Figure 13(a) demonstrates the co-registered point cloud combining the optical and two TomoSAR point clouds. The figure shows the identical area as that in Figure 11 and Figure 12. Successful co-registration can be confirmed by seeing the correct location of the façade points in Figure 13(b) which shows a close-up view of the matched point cloud with different colors representing different point clouds. The ground points in Figure 13(b) were removed for a better visualization.



(a)



(b)

Figure 13. (a) The fused point cloud combining the optical point cloud and two TomoSAR point clouds from ascending and descending viewing angle. The height is color-coded. The colorbar and the area are identical to those in Figure 11 and Figure 12, and (b) a close up view of the co-registered point cloud in Berlin Potsdamer Platz, where green, red, blue represent the points from optical, ascending and descending TomoSAR point cloud, respectively. The ground points were removed for a better visualization.

As no ground truth of the co-registration of the optical and TomoSAR point clouds is available, we used a simulation to assess the performance of the robust AICP. A LiDAR point cloud of a building in Berlin about 10,000 points with centimeter accuracy was used as the reference point cloud, i.e. optical point cloud. We simulated non-ergodic noise for the reference point cloud with SNR distributed from -10 to 10dB, which well simulated the noise in the TomoSAR point cloud. The noisy LiDAR point cloud is then translated to simulate the target point cloud, i.e. TomoSAR point cloud. Figure 14 shows the original and the noisy LiDAR point clouds.

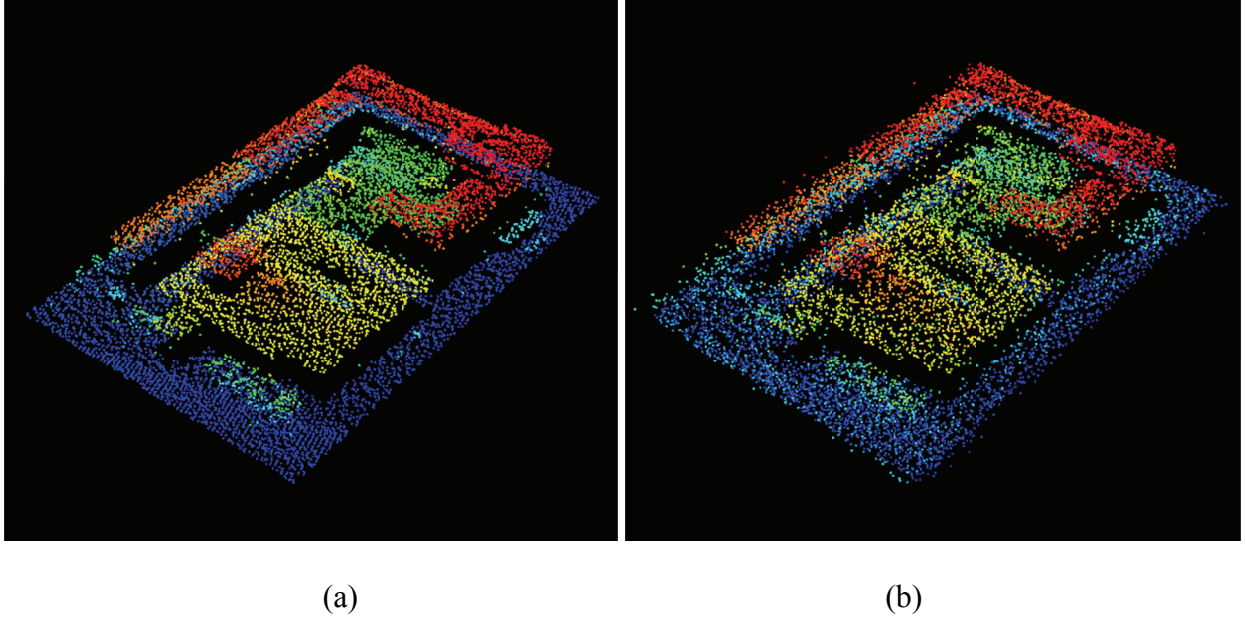


Figure 14. (a) original LiDAR point cloud with centimeter accuracy, and (b) simulated TomoSAR point cloud by adding non-ergodic noise to the LiDAR point cloud.

Figure 15 compares the translation error of the original ICP, the AICP, and AICP with M-estimate of covariance matrix (MAICP) at each iteration. All three algorithms converge. Comparison shows that MAIPC outperforms the other methods in terms of co-registration accuracy and convergence rate. In theory, AICP should also outperform ICP. But the incorrect covariance matrix estimation due to non-ergodic noise leads to its worse performance than the original ICP. This implies a robust estimation of the covariance matrices is necessary in such case. The same experiment was also performed on several other typical benchmark point clouds, such as the Stanford bunny. Consistent results were obtained.

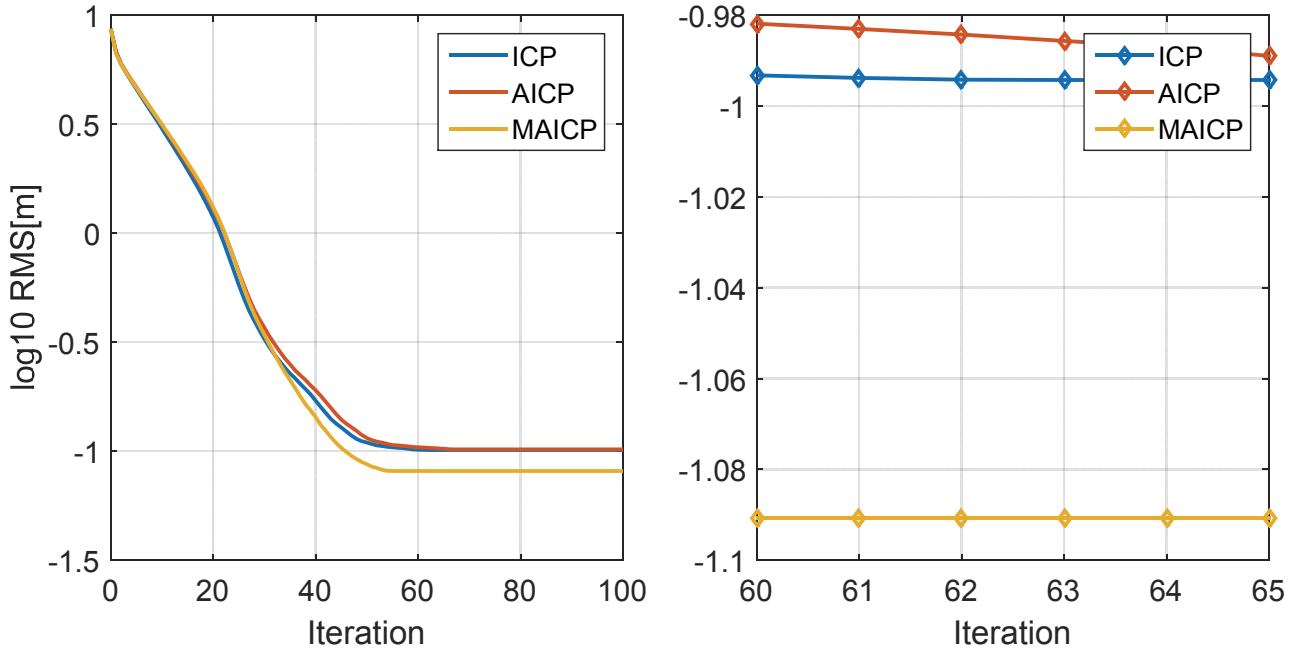
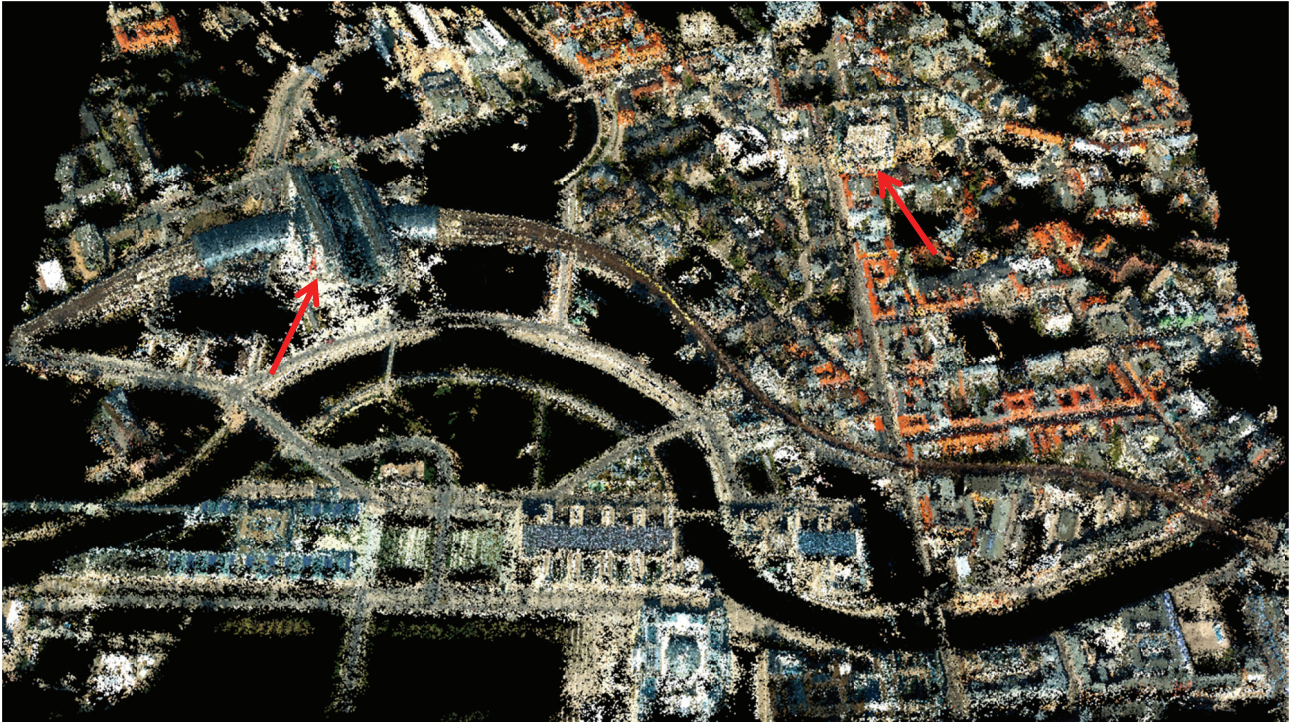


Figure 15. Comparison of the translation error of original ICP, anisotropic ICP, and anisotropic ICP with robust covariance matrix estimate (M-estimate) at each iteration of the algorithms. The scale is \log_{10} [m]. The right subfigure is a close up view of iteration 60 to 65.

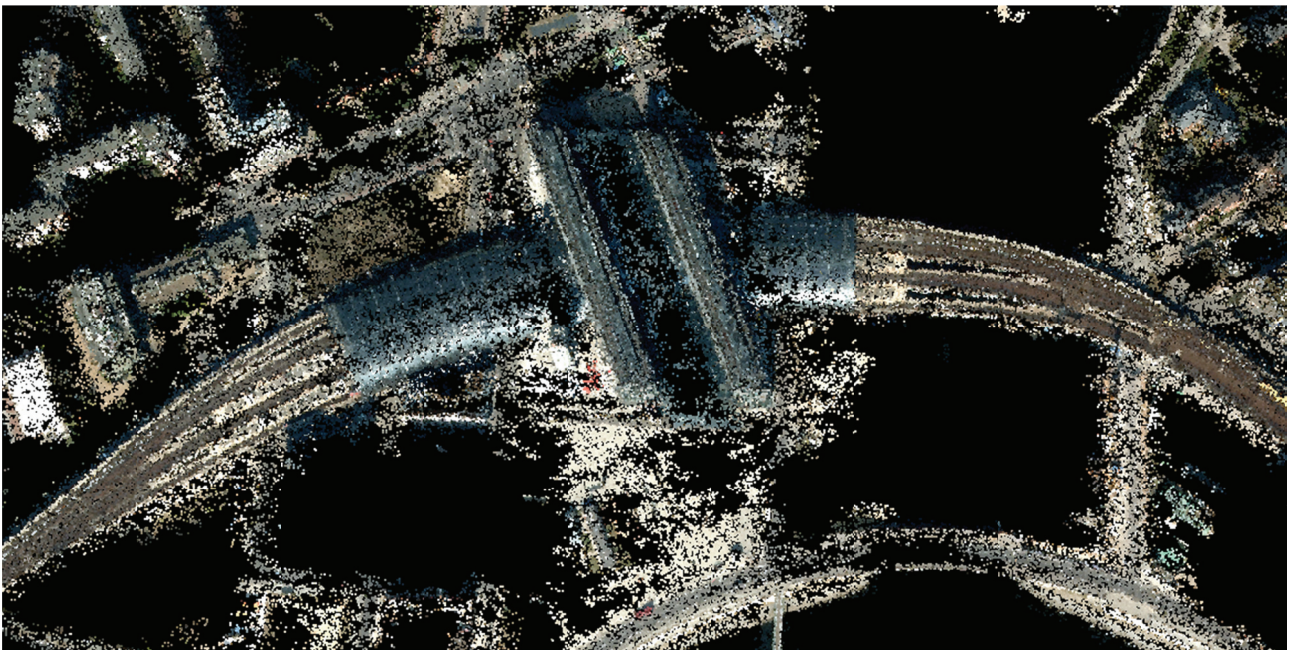
6.5 TomoSAR Point Cloud Texturing

By applying the texturing algorithm, Figure 16 shows the first meter-resolution spaceborne TomoSAR point cloud of an urban area textured with optical color. The subfigure (b) and (c) are zoom in to the Berlin central station and the high-rise building of Universitätsmedizin Berlin which are indicated by the red arrows in subfigure (a). The black parts are where no TomoSAR point was reconstructed. The proposed framework successfully projects optical images to urban TomoSAR point cloud, including points on large façades, as well as points on ground objects.

On the other hand, it is also possible to project the 2-D optical image to the 2-D SAR image geometry. Figure 17 shows the SAR amplitude image and the optical image projected in SAR range-azimuth geometry. This is also the first high resolution urban optical image projected in SAR geometry. The Berlin central station and the hospital are also indicated by the red arrows in Figure 17.



(a)



(b)



(c)

Figure 16. (a) First TomoSAR point cloud textured with optical image, and (b) zoom in to the Berlin central station, and (c) zoom in to the high-rise building of Universitätsmedizin Berlin. The coordinate is 3-D UTM. The proposed framework successfully projects optical images to urban TomoSAR point cloud, including points on large façades, as well as points on ground objects.

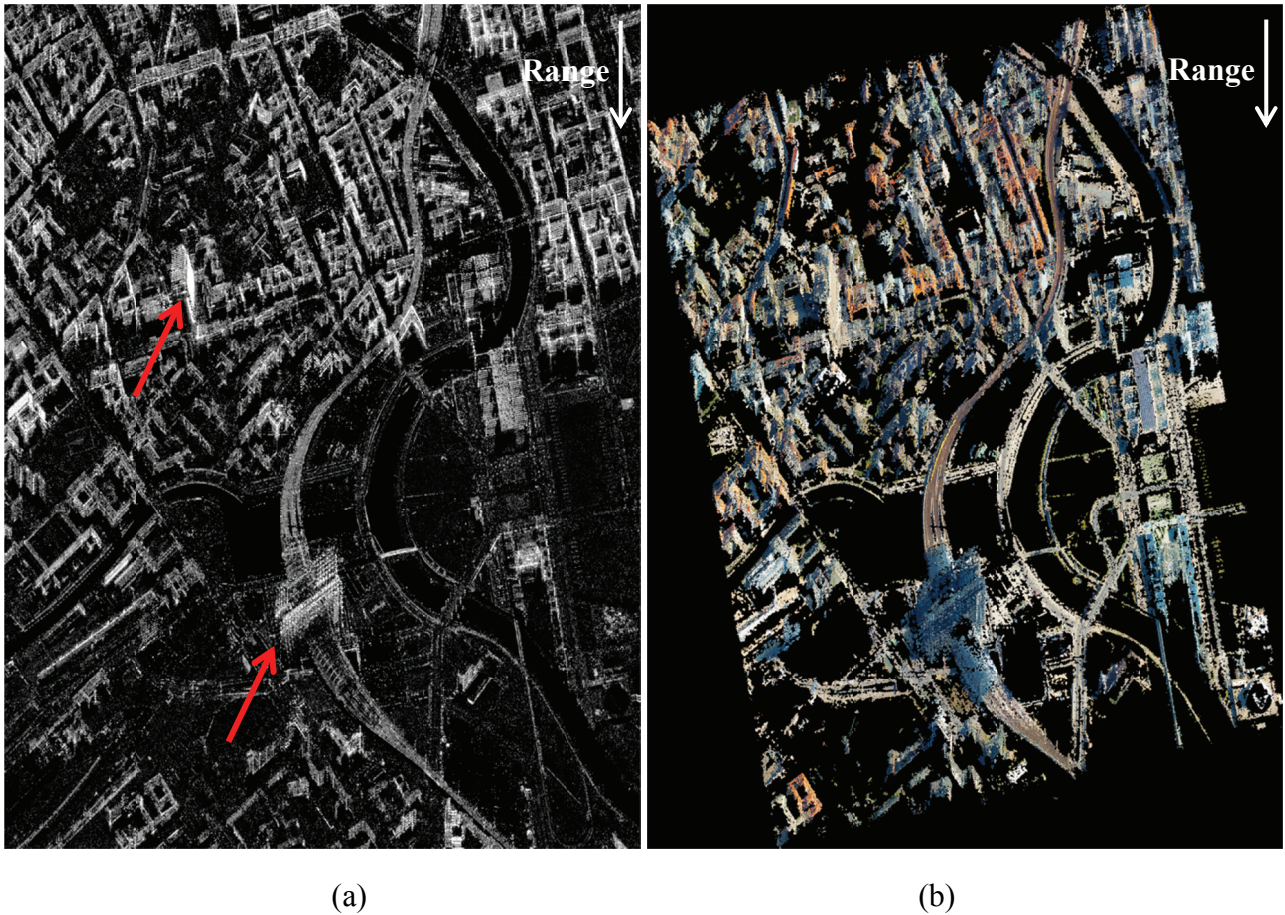


Figure 17. (a) the SAR amplitude image, and (b) the optical image in SAR image geometry. Both images are in SAR azimuth and range coordinate. The red arrows indicate the locations of Berlin central station and the high-rise building of Universitätsmedizin Berlin.

6.6 Optical Image Classification

The optical images were classified patch-wisely using the Bag of Words (BoW) [57] method which is a well-known technique in the computer vision community. As this is not the focus of this article, the readers are kindly referred to the original literature for more information. For the deformation analysis in this paper, we classified the railway and the river class, since we found interesting deformation pattern on the railway and bridges. Classification on large façades is also a potential application, which of course requires oblique optical images.

The classification is supervised. Training patches were manually selected for each class. The classification is done patch-wisely in the large aerial image. The feature used is simply the RGB value in a 3×3 sliding window in the patch. The classifier is a linear SVM [58] implemented in an open source library VLFeat [59]. Figure 18(a) shows the classification result, where red is river, green is railway, and blue is bridge. The bridge label is derived by filling the gap between the river segments. Small clusters do appear as false alarm. But they can be easily removed by post-processing. These labels can be projected to the SAR image or the TomoSAR point cloud, in order to perform further analysis. For example, Figure 18(b) is the railway label (after post-processing, explained in section 6.7) projected to the SAR image geometry identical to that of Figure 17.

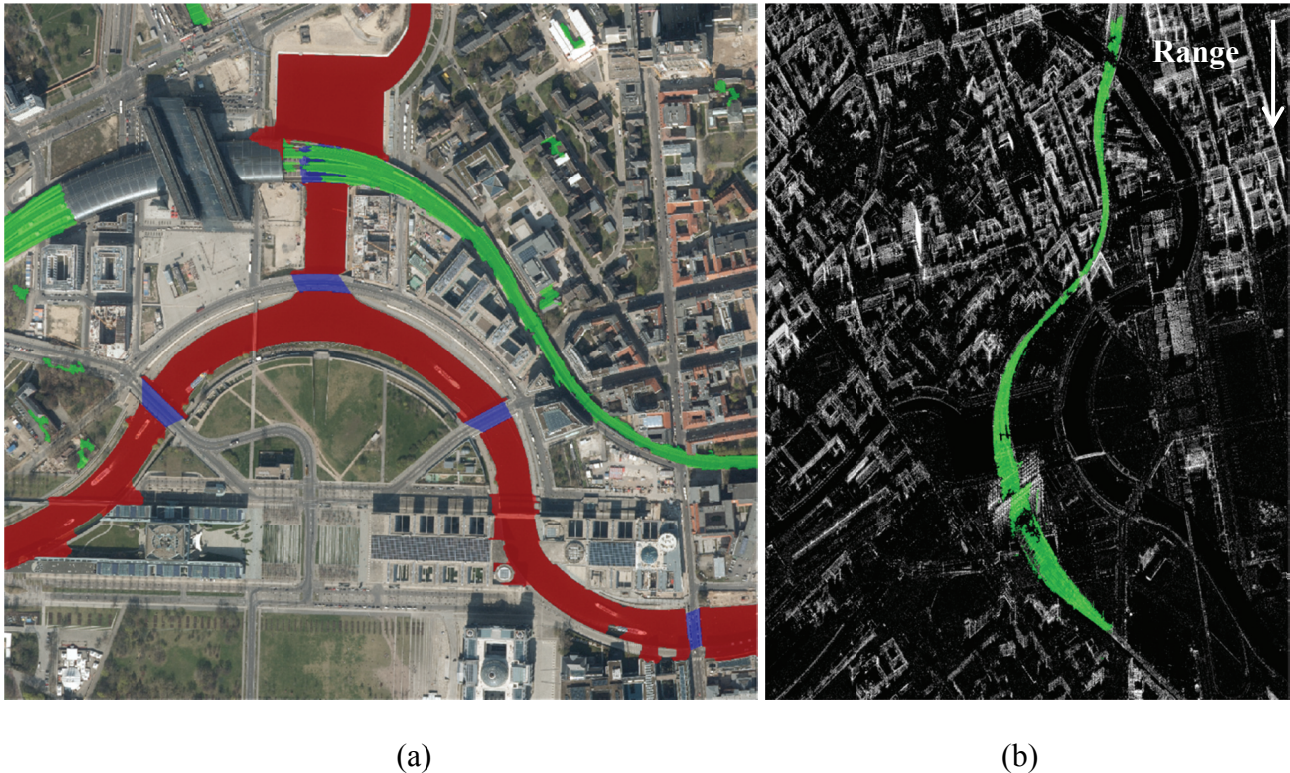


Figure 18. (a) River (red), railway (green) and bridge (blue) classified using the BoW method. The blue label is derived by filling the gap between the river segments, which indicates the bridge locations. Some false alarm appeared as small clusters. They can be removed using post-processing. (b) The railway label (after post-processing explained in section 6.7) projected to SAR image geometry which is identical to that in Figure 17.

6.7 Railway Monitoring

Based on the classification, the corresponding points in the TomoSAR point cloud can be extracted by projecting the classification label to the point cloud. A smooth spline function was fitted to the east-north coordinates of the railway points to connect separated segments. For example, the railway label is disconnected due to the presence of the Berlin central station. Figure 19(a) shows the extracted continuous railway points overlaid on the optical image. The color shows the amplitude of seasonal motion.

This motion is mostly caused by the thermal dilation of the steel railway because of the seasonal temperature change. It can be observed that the railway deformation experiences certain periodic discontinuity. It is suspected to be caused by the mechanical joints between different railway segments.

For monitoring these discontinuities, the raw TomoSAR estimates are relatively noisy due to the relative low SNR of some scatterers. To filter out the noise, a special prior was employed, i.e. the thermal dilation of a steel beam is linearly proportional to its length at its first approximation [60]. That is to say, the railway deformation parameters is piecewise linear, and the scatterers on the same cross-section of the railway undergo similar deformation. Therefore, the deformation estimates can be filtered by minimizing its total generalized variation along the railway direction (with second order derivative which favors piecewise linear structure):

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \left\{ \frac{1}{2} \|\mathbf{f} - \mathbf{a}\|_2^2 + \lambda \|\Delta \mathbf{f}\|_1 \right\} \quad (16)$$

where \mathbf{a} is the deformation estimates along the 1-D railway direction (i.e. unfold the railway to a straight line), \mathbf{f} is the filtered version of \mathbf{a} , and $\Delta \mathbf{f}$ is the second order derivative of \mathbf{f} . As shown

in literatures of total generalized variation [61], [62], the L_1 norm of second order derivative is convex and lower semi-continuous, one can solve it using convex optimization solvers.

The filtered deformation estimates are shown in Figure 19(b), in comparison with the unfiltered estimates Figure 19(a). The noise has been greatly reduced while the discontinuity is well preserved. To have a quantitative comparison, Figure 20(a) plots the original (in blue) and the denoised (in red) deformation estimates as a function of the 1-D railway direction. The original estimates have a standard deviation about 1~2 mm which concealed some edges where the magnitude of the deformation is low. The filtering successfully reconstructs the edges.

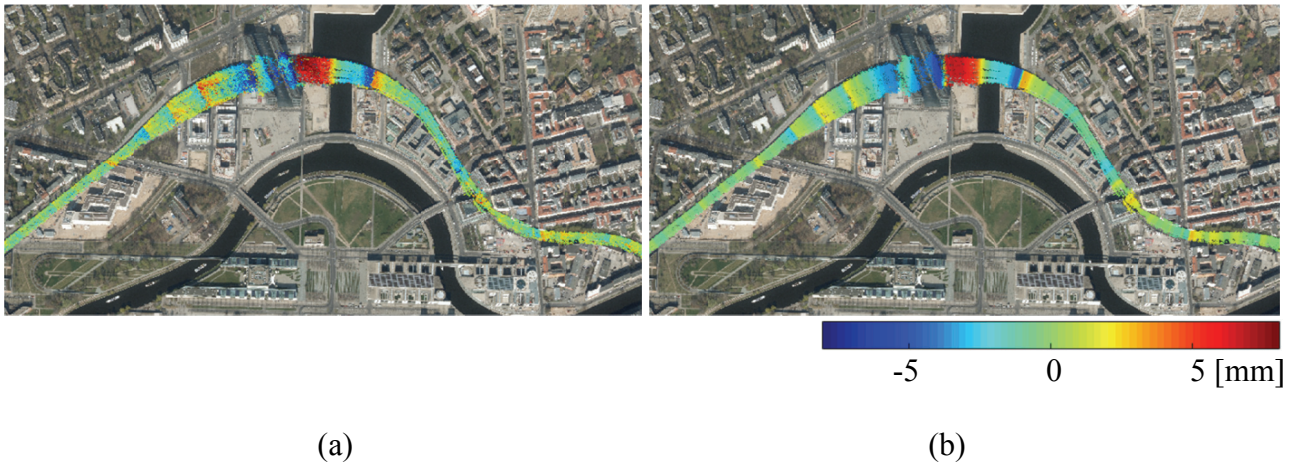
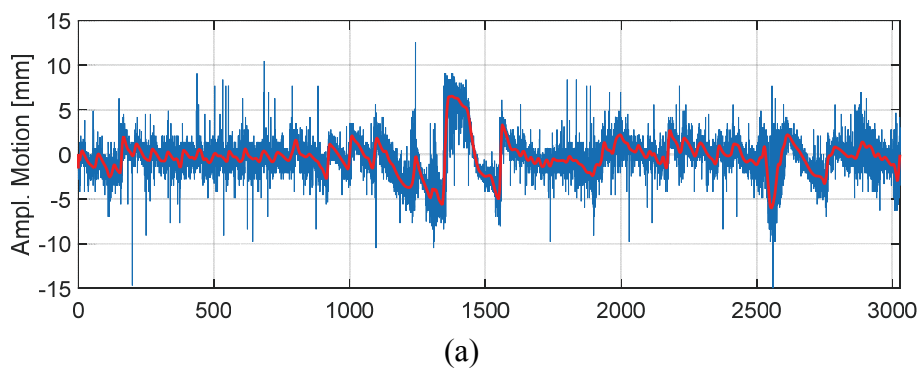


Figure 19. (a) the railway points extracted from the original TomoSAR point cloud, and (b) the railway points filtered using total generalized variation. The color shows the amplitude of seasonal motion due to the thermal expansion of the railway.



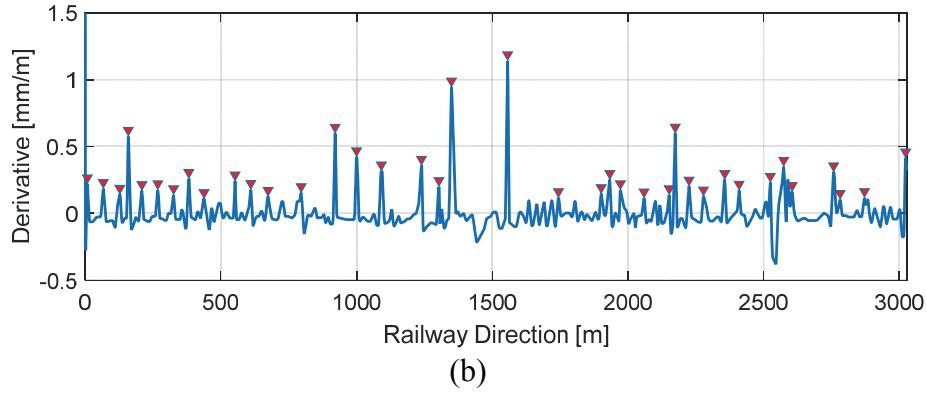


Figure 20. (a) the original and the filtered estimates of amplitude of seasonal deformation in blue and red, respectively, and (b) peaks detected in the derivative of the filtered deformation along the railway direction.

By detecting the peaks in the derivative of the deformation function, the location of the discontinuities can be detected. Constraint was put on the minimum distance between two peaks representing the minimum length of a railway segment. The detected peaks in the deformation's derivative can be seen in Figure 20(b). The positions of the peaks on the railway are shown as the green dots in Figure 21. Each green dot represents the midpoint of its railway cross-section. In the middle subfigures of Figure 21, the close up view of the two joints in the optical image is provided. As the optical image has limited resolution, we also provide a higher resolution one (7cm ground spacing) in Figure 22. It can be clearly observed that the railway joint shown up as dark lines in the optical image.

However, further investigation shows that such large mechanical joints only appear close to railway stations. Visual inspection of the high resolution image of other parts of the railway does not find obvious mechanical joints. Therefore, a more general cause of the discontinuities in the deformation still requires further investigation.



Figure 21. Upper: the midpoint of the detected railway joint cross-section marked in green, and lower: close up view of the railway joints. The background optical image has a ground spacing of 20cm.



Figure 22. A higher resolution image (7cm ground spacing) of the lower left subfigure of Figure 21.

6.8 Bridges Monitoring

By analyzing the gaps of the river segmentation and assuming the gaps are caused by bridges, the bridges' positions can be detected. Without going into too much detail, the extracted bridges points overlaid on the optical image are shown in Figure 23 where the color represents the amplitude of seasonal deformation. The upper bridge belongs to a segment of the railway which is known to have thermal expansion. The middle bridge undergoes a 5mm seasonal motion on its west end and 2mm at the east end. This suggests a more rigid connection of the bridge to the foundation at its east end. The two lower bridges are stable according to the motion estimates.

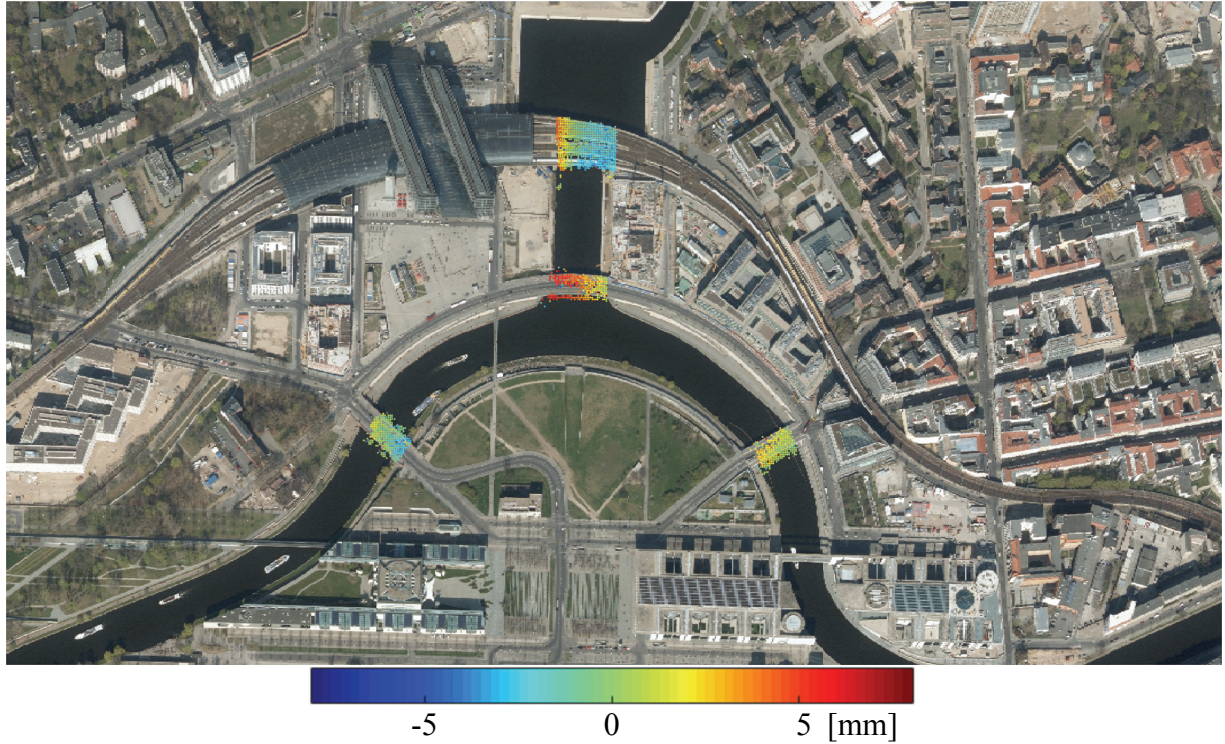


Figure 23. The amplitude of seasonal motion of the bridges extracted from the TomoSAR point cloud overlaid on the optical image.

CONCLUSION AND OUTLOOK

This paper presents the first joint semantic analysis of optical images and meter-resolution spaceborne InSAR point cloud in urban area. Through a 3-D geometric fusion of optical images and InSAR point cloud, the attributes from the optical image, e.g. color, semantic labels, can be projected to the InSAR point cloud. We presented the first urban “SARptical” image which is the optical image projected to SAR image geometry.

Experiments show that a robust estimation of point covariance matrix is mandatory in applying anisotropic ICP algorithm to TomoSAR point cloud. Example on railway monitoring discovers the periodic deformation pattern along the railway in Berlin is caused by the mechanical joints between railway segments only for several locations close to railway stations. No visible mechanical joint is identified for other location, which leads to some further investigation of the cause of deformation.

This work also opens some new perspectives in high resolution urban remote sensing. For example, *systematic monitoring of large façades*, which requires high resolution oblique optical images and the development of façade classification algorithm; and *joint classification of high resolution optical image and InSAR point cloud/SAR image*, which can utilize InSAR attributes such as 3-D position and deformation parameters as additional features in the classification. For all these methods, the main challenge remains at the positioning accuracy of spaceborne InSAR point cloud, which is in the order of 1 to 10 m for TerraSAR-X with its three imaging modes. Future work on improving the positioning accuracy of TomoSAR, e.g. by exploiting joint sparsity [63] or object based TomoSAR inversion [64], is the key in the joint analysis of high resolution urban optical and SAR images.

ACKNOWLEDGEMENT

This work was supported by the Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018, www.sipeo.bgu.tum.de), International Graduate School of Science and Engineering, Technische Universität München (Project 6.08: “4D City”) and the German Aerospace Center (DLR, Förderkennzeichen 50EE1417). The authors would like to thank Dr. H. Hirschmüller of DLR-RM for providing the optical data, as well as the reviewers for their valuable suggestions.

REFERENCES

- [1] R. Bamler, M. Eineder, N. Adam, X. Zhu, and S. Gernhardt, “Interferometric Potential of High Resolution Spaceborne SAR,” *Photogramm. - Fernerkund. - Geoinformation*, vol. 2009, no. 5, pp. 407–419, Nov. 2009.
- [2] U. Soergel, Ed., *Radar Remote Sensing of Urban Areas*, vol. 15. Dordrecht: Springer Netherlands, 2010.
- [3] Y. Wang and X. X. Zhu, “InSAR Forensics: Tracing InSAR Scatterers in High Resolution Optical Image,” presented at the Fringe 2015, 2015.
- [4] Y. Wang and X. X. Zhu, “Semantic Interpretation of InSAR Point Cloud,” in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015*, Milan, Italy, 2015.
- [5] C. Frueh, R. Sammon, and A. Zakhor, “Automated texture mapping of 3D city models with oblique aerial imagery,” in *2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings*, 2004, pp. 396–403.

- [6] F. Lombardini, "Differential tomography: a new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.
- [7] G. Fornaro, D. Reale, and F. Serafino, "Four-Dimensional SAR Imaging for Height Estimation and Monitoring of Single and Double Scatterers," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 224–237, Jan. 2009.
- [8] X. Zhu and R. Bamler, "Very High Resolution Spaceborne SAR Tomography in Urban Environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, 2010.
- [9] X. Zhu and R. Bamler, "Tomographic SAR Inversion by L1-Norm Regularization -- The Compressive Sensing Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, 2010.
- [10] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.
- [11] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.
- [12] X. X. Zhu and R. Bamler, "Let's Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, 2011.
- [13] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, p. 105, Oct. 2011.
- [14] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, and others, "Detailed real-time urban 3d reconstruction from video," *Int. J. Comput. Vis.*, vol. 78, no. 2–3, pp. 143–167, 2008.
- [15] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual Modeling with a Hand-Held Camera," *Int J Comput Vis.*, vol. 59, no. 3, pp. 207–232, Sep. 2004.
- [16] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the World from Internet Photo Collections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 189–210, Nov. 2008.
- [17] C. Wu, "Towards linear-time incremental structure from motion," in *3D Vision-3DV 2013, 2013 International Conference on*, 2013, pp. 127–134.
- [18] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [19] M. Pollefeys, R. Koch, and L. Van Gool, "Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters," *Int. J. Comput. Vis.*, vol. 32, no. 1, pp. 7–25, 1999.
- [20] M. Ali, D. Clausi, and others, "Automatic registration of SAR and visible band remote sensing images," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*, 2002, vol. 3, pp. 1331–1333.
- [21] H. Cheng, S. Zheng, Q. Yu, J. Tian, and J. Liu, "Matching of SAR images and optical images based on edge feature extracted via SVM," in *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, 2004, vol. 2, pp. 930–933.
- [22] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of Optical and SAR Satellite Images by Exploring the Spatial Relationship of the Improved SIFT," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 657–661, Jul. 2013.
- [23] T. D. Hong and R. A. Schowengerdt, "A robust technique for precise registration of radar and optical satellite images," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 5, pp. 585–593, 2005.
- [24] H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *Image Process. IEEE Trans. On*, vol. 4, no. 3, pp. 320–334, 1995.
- [25] G. Palubinskas and P. Reinartz, "Template based matching of optical and SAR Imagery," in *Joint Urban Remote Sensing Event (JURSE) 2015*, Lausanne, Switzerland, 2015, pp. 1–4.

- [26] S. Auer, C. Gisinger, and J. Tao, "Characterization of Facade Regularities in High-Resolution SAR Images," *Geosci. Remote Sens. IEEE Trans. On*, vol. 53, no. 5, pp. 2727–2737, May 2015.
- [27] J. Tao, S. Auer, G. Palubinskas, P. Reinartz, and R. Bamler, "Automatic SAR Simulation Technique for Object Identification in Complex Urban Scenarios," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 3, pp. 994–1003, Mar. 2014.
- [28] L. Schack, U. Soergel, and C. Heipke, "Persistent Scatterer Aided Facade Lattice Extraction in Single Airborne Optical Oblique Images," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. II-3/W4, pp. 197–205, Mar. 2015.
- [29] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.
- [30] S. Gernhardt and R. Bamler, "Deformation monitoring of single buildings using meter-resolution SAR data in PSI," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 68–79, Sep. 2012.
- [31] Y. Wang and X. Zhu, "Automatic Feature-based Geometric Fusion of Multi-view TomoSAR Point Clouds in Urban Area," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 3, pp. 953 – 965, 2015.
- [32] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 119–152, 1994.
- [33] R. S. J. Estépar, A. Brun, and C.-F. Westin, "Robust generalized total least squares iterative closest point registration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*, Springer, 2004, pp. 234–241.
- [34] L. Maier-Hein, T. R. dos Santos, A. M. Franz, and H.-P. Meinzer, "Iterative Closest Point Algorithm in the Presence of Anisotropic Noise.," *Bildverarb. Für Med.*, vol. 2010, pp. 231–235, 2010.
- [35] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, "Imaging Geodesy - Toward Centimeter-Level Ranging Accuracy With TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 661–671, Feb. 2011.
- [36] X. Cong, U. Balss, M. Eineder, and T. Fritz, "Imaging Geodesy — Centimeter-Level Ranging Accuracy With TerraSAR-X: An Update," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 948–952, Sep. 2012.
- [37] M. Eineder, U. Balss, C. Gisinger, S. Hackel, X. Cong, F. G. Ulmer, and T. Fritz, "TerraSAR-X pixel localization accuracy: Approaching the centimeter level," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 2014, pp. 2669–2670.
- [38] I. Sobel, "An Isotropic 3x3 Image Gradient Operator," *Present. Stanf. AI Proj.* 1968, 1968.
- [39] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 1981.
- [40] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.
- [41] Y. Wang and X. X. Zhu, "Robust Estimators for Multipass SAR Interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 968–980, Feb. 2016.
- [42] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," in *Statistical Signal Processing, 2003 IEEE Workshop on*, 2003, pp. 462–465.
- [43] M. Shahzad and X. X. Zhu, "Robust Reconstruction of Building Facades for Large Areas Using Spaceborne TomoSAR Point Clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 752–769, Feb. 2015.
- [44] X. Zhu and M. Shahzad, "Façade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.

- [45] M. Shahzad and X. X. Zhu, "Automatic Detection and Reconstruction of 2-D/3-D Building Shapes From Spaceborne TomoSAR Point Clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1292–1310, Mar. 2016.
- [46] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," *ACM Trans. Graph. TOG*, vol. 26, no. 3, p. 24, 2007.
- [47] R. Mehra, P. Tripathi, A. Sheffer, and N. J. Mitra, "Visibility of noisy point cloud data," *Comput. Graph.*, vol. 34, no. 3, pp. 219–230, 2010.
- [48] F. Rodriguez Gonzalez, N. Adam, A. Parizzi, and R. Brcic, "The Integrated Wide Area Processor (IWAP): A Processor for Wide Area Persistent Scatterer Interferometry," presented at the ESA Living Planet Symposium, Edinburgh, UK, 2013.
- [49] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *ISPRS Workshop High Resolution Mapping from Space, Hannover, Germany*, 2003, vol. 2003, p. 6.
- [50] X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's Tomographic SAR Processing System," in *Urban Remote Sensing Event (JURSE), 2013 Joint*, 2013, pp. 159–162.
- [51] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring: A Sparse and Nonlinear Tour*, vol. 666. Deutsche Geodätische Kommission, 2011.
- [52] X. Zhu and R. Bamler, "Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, 2012.
- [53] Y. Wang, X. Zhu, and R. Bamler, "An Efficient Tomographic Inversion Approach for Urban Mapping Using Meter Resolution SAR Image Stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1250–1254, 2014.
- [54] "<https://www.pix4d.com/>." 2015.
- [55] X. X. Zhu, S. Montazeri, C. Gisinger, R. F. Hanssen, and R. Bamler, "Geodetic SAR Tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 18–35, Jan. 2016.
- [56] C. Gisinger, U. Balss, R. Pail, X. X. Zhu, S. Montazeri, S. Gernhardt, and M. Eineder, "Precise Three-Dimensional Stereo Localization of Corner Reflectors and Persistent Scatterers With TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1782–1802, Apr. 2015.
- [57] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, 2004, vol. 1, pp. 1–2.
- [58] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [59] B. F. Andrea Vedaldi, "VLFeat: an open and portable library of computer vision algorithms," in *Proceedings of the 18th International Conference on Multimedia 2010*, Firenze, Italy, 2010, pp. 1469–1472.
- [60] A. D. Kerr, "Analysis of thermal track buckling in the lateral plane," *Acta Mech.*, vol. 30, no. 1–2, pp. 17–50, 1978.
- [61] K. Bredies, K. Kunisch, and T. Pock, "Total Generalized Variation," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, Jan. 2010.
- [62] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magn. Reson. Med.*, vol. 65, no. 2, pp. 480–491, 2011.
- [63] X. Zhu, N. Ge, and M. Shahzad, "Joint Sparsity in SAR Tomography for Urban Mapping," *IEEE J. Sel. Top. Signal Process.*, vol. PP, no. 99, pp. 1–12, 2015.
- [64] J. Kang, Y. Wang, M. Köner, and X. X. Zhu, "Object-based InSAR Deformation Reconstruction with Application to Bridge Monitoring," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2016*, Beijing, China.